

क्या एआई को नैतिक नियमों में बांधा जा सकता है?



मशीनें और आर्टिफिशियल इंटेलिजेंस, एआई या कृत्रिम बुद्धिमत्ता का मानव-जीवन के निर्णय लेने, विशेषरूप से प्रशासन में तेजी से प्रभाव बढ़ता जा रहा है। इसे देखते हुए बहुत से देशों ने एआई संबंधी नियमन तैयार किए हैं। प्रश्न यह उठता है कि मशीनों और एआई को ऐसे नियमों के जरि, क्या नैतिकता में बांधा जा सकता है? पहले एआई से संबंधित नैतिक चुनौतियों पर कुछ बिंदु -

- रोबोटिक्स की दुनिया में नैतिकता को नियंत्रण में रखने के लिए आइजैक असिमोव ने 'थ्री लॉज ऑफ रोबोटिक्स' की रचना की। लेकिन ऐसा देखा गया कि इसके नियम अक्सर अप्रत्याशित और विरोधाभासी परिणामों की ओर ले जाते हैं। रोबोट हो या एआई आधारित प्रशासनिक निर्णय क्षमता, इनमें मानवीय नैतिकता को एल्गोरिदम में प्रोग्राम करना अत्यंत जटिल है।
- जेम्स मूर ने भी शतरंज की चालों को प्रोग्राम करने की तुलना में नैतिकता को एआई में प्रोग्राम करना बहुत जटिल बताया है। उनका कहना है कि नैतिकता कुछ अपरिभाषित चालों के साथ एक जटिल तरीके से वैध काम करती है।

मूर के 2006 के वर्गीकरण के अनुसार मशीनी नैतिकता के चार वर्ग -

- नैतिक प्रभाव वाले एजेंट- रोबोट ये नैतिकता संबंधी निर्णय नहीं ले सकते, लेकिन कुछ विकल्प प्रस्तुत करते हैं।
- अंतर्निहित नैतिक एजेंट-विमान का ऑटोपायलट सिस्टम यह नैतिकता तय किए बिना निर्धारित नियमों का पालन करता है।

- स्पष्ट नैतिक एजेंट- यह निर्धारित नियमों के परे जा सकता है। उदाहरण के लिए, वित्तीय निवेश को संतुलित करने वाला एआई।
- पूर्ण नैतिक एजेंट- यह उचित स्पष्टीकरण सहित नैतिक निर्णय लेने में सक्षम है। एक वयस्क मनुष्य की तरह उन्नत एआई भी ऐसा हो सकता है ऐसा एआई अभी तक बनाया नहीं गया है।

इन चार वर्गीकरण से संबंधित अनेक पत्र प्रकाशित हुए हैं, जिनमें अंतिम निष्कर्ष यही निकलता है कि तीसरे और चौथे वर्ग वाले एजेंट की रचना मुश्किल है। अगर मानवीय मॉडल पर आधारित नैतिकता की प्रोग्रामिंग करना संभव भी हो, तो क्या गारंटी है कि ऐसा एआई मानव जैसे अनैतिक व्यवहार में संलग्न न हो। इससे पता चलता है कि मानवीय नैतिक पतन, मशीन नैतिकता का भी हिस्सा हो सकता है।

इस स्थिति में किसे जिम्मेदार ठहराया जाएगा? एआई, सिस्टम को, उसके डेवलपर को या उस अधिकारी को जो एआई डेटा पर निर्भर था? ये वे सवाल हैं, जिनका उत्तर सरकारों को ढूंढना है।

'द हिंदू' में प्रकाशित आदित्य सिन्हा के लेख पर आधारित। 24 अगस्त, 2023



AFEIAS